Additional reflections on Putnam, Wright and Brains in Vats

Putnam's argument against the sceptical Brain-in-a-Vat hypothesis continues to intrigue. I argue in what follows that the argument refutes a particular kind of sceptic and make a proposal about its more general significance. To appreciate the soundness of the argument, I explain, we need to appreciate that the sceptic's contention is that I cannot know that I am not a brain in a vat *even if* I am not. This is why in response to the sceptic it is legitimate to make a transition from knowing that a sentence is true to knowing the truth it expresses, which is the crucial move in the argument.

Following Crispin Wright (1992) we can express Putnam's Brain-in-a-Vat argument against the sceptic (see his (1981), chapter 1) as follows:

(A) 'Snow is white' in my language means that snow is white

(B) If I am a BIV then 'Snow is white' in my language does not mean that snow is white

(C) I am not a BIV [from (A) and (B)].

(Putnam, and following him Crispin Wright, has 'I am a BIV' where 'Snow is white' ('S is W' hereafter) occurs in the formulation in the text. What difference does this make? There are three points to be made. (1) Replacement of 'S is W' by 'I am a BIV' would make no difference to the *soundness* of the argument. (2) It would, however, guarantee the argument's availability for first-personal rehearsal by any philosophically knowledgeable reader of this paper, independently of the contingencies of his linguistic resources - whether, that is, he be an Alaskan Eskimo or Saharan Arab. (Saharan Arabs are hereby permitted to replace 'S is W' throughout the text by 'Palm trees are green'.) (3) Replacement of 'S is W' by 'I am a brain in a vat' would enable the conclusion of the argument to be expressed as the claim that the sceptic's position is *self-refuting* (as, indeed, Putnam does express it), or, as Crispin Wright puts it, *absaud* (sic). But, as Wright in effect notes towards the end of his paper, so long as the argument refutes the sceptic, that is enough, self-refutation is merely the icing on the cake.)

(A)-(C) is an argument for the conclusion that I am not a BIV. The sceptic's claim is that I *cannot know* that I am not – I cannot know that I am not *even if* I am not (of course, I cannot know that I am not if I am). I cannot know that I am not even if I am not since I cannot

know this *a priori*, i.e., without empirical investigation additional to any required to acquire the concepts in the proposition known, and if I cannot know it *a priori* I cannot know it at all. It will therefore suffice to respond to the sceptic to establish the *conditional* conclusion that *if* I am not a BIV, I can (somehow) know that I am not, or to establish the conclusion that I can know that I am a not a BIV on the *assumption* that I am not. To do this is not to argue for the denial of the sceptic's conclusion, which is that I do not know that I am not a brain in a vat – and so is identical with the conclusion that can be drawn by the lunatic who actually thinks that I am a brain in a vat. But it is to argue that the sceptic (unlike the lunatic) has no basis on which to assert his conclusion.

To get an argument for the conclusion that I can know that I am not a brain in a vat we can therefore restate Putnam's argument as follows:

(1) I can know that 'S is W' means in my language that S is W

(2) I can know that if I am a BIV then 'S is W' in my language does not mean that S is W

(3) I can know that I am not a BIV [from (1) and (2)]

Since anyone who can know both (A) and (B) above must be capable of knowing (C) (anyone who can know that [Q and if P then not-Q] can know that not-P), the only way to deny the validity of this argument is to say that a situation is possible in which someone can know that (A) and can know that (B) without being capable of knowing [(A) and (B)]. The sceptic is best advised therefore rather to query the truth of the premises or the right of his opponent to assert them.

The sceptic may object to the first premiss of this argument on the ground that, when the knowledge that 'S is W' means that S is W in my language is distinguished from the knowledge that the *sentence* '"S is W" means that S is W' expresses a (some) truth in my language, it can be seen to be question-begging to assume it in the context of this debate (for the BIV does not know that 'S is W' means that S is W in *his* language, since it does not, though he does know that the sentence '"S is W" means that S is W' expresses a truth in his language).[1]

However, (1) is dispensable as a premiss, as the following modification of the

[1] Wright notes this worry on p.75 of his paper.

argument shows (in which, for good measure (2) has also been derived from two supplementary premisses in order to make its content clearer):

(1a) I can know that (whatever it means) the sentence '"S is W" in my language means that S is W' is true

(1b) I understand the sentence '"S is W" in my language means that S is W'

(1c) '"S is W" in my language means that S is W' means in my language that 'S is W' in my language means that S is W

(1) I can know that 'S is W' in my language means that S is W [from (1a), (1b) and (1c)]

(2a) I can know that if whatever is necessary for its being the case that 'S is W' means that S is W in my language is not the case, then 'S is W' does not mean that S is W in my language

(2b) I can know that a necessary condition of its being the case that 'S is W' means that S is W in my language is that I am not a BIV

(2) I can know that if I am a BIV then 'S is W' in my language does not means that S is W [from (2a) and (2b)]

(3) I can know that I am not a BIV [from (1) and (2)]

The argument from (1a), (1b) and (1c) to (1) is of the form:

(1af) I can know that S is true

(1bf) I understand the sentence S

(1cf) The sentence S means in my language that p

(1f) I can know that p

This is a valid form of argument, as is the rest of the argument from (1) on to (3). (More cautiously, there is a reading on which this is a valid form of argument.) Moreover (1a), (1b) and (1c) are all *unambiguously* true)

How can the sceptic respond? He must question at least one of the premises if he is to resist. Which one(s)?

It would not be relevant for the sceptic to object to (1b) or (1c).

Re (1b): it is part of the BIV scenario that *the BIV* understands the sentence '"S is W" in my language means that S is W' – in his own way, of course. And I am no worse off in respect of understanding than the BIV, so I understand that sentence too – in my own way, of

course. To put this point another way. The sceptic Putnam's argument is directed against is not Kripke's (inappropriately named) 'sceptic about meaning', who denies the possibility of meaning and understanding and so Putnam's sceptic has no ground for denying (1b). Of course, he *can* do so in order to block the argument if he can see no other way of resisting and yet wishes to retain his scepticism, but if the only recourse of Putnam's sceptic is to embrace Kripkean meaning scepticism that in itself is a significant and surprising conclusion.

Re (1c): this simply says *what* the quoted sentence in fact means; it does not say anything about anyone's knowing that it means this and so it would be irrelevant for the sceptic to question it. Here it is important to recall that the sceptic must say that I cannot know that I am not a BIV *even if* I am not one – so, in particular, he must say that I cannot know that I am not a BIV even if the sentence '"S is W" means in my language that S is W' means in my language that 'S is W' means in my language that S is W (which it does since I am not a BIV and speak ordinary English, but would not if I were and did not). In responding to him (1c) can therefore be legitimately assumed. Crucially, I do not have to be in a position to *assert* it.

The sceptic might try, rather desperately, I think, to stick at (1a). But what (1a) says of me is also true of the BIV (it is true of me if and only if it is true of the BIV), so again it would not be relevant for the sceptic to object. However, in fact (1a), like (1), seems to be dispensable as a premiss, for the following argument seems to be sound:

(1b) I understand the sentence '"S is W" my language means that S is W'

(1c) '"S is W" in my language means that S is W' means in my language that 'S is W' in my language means that S is W

(1e) I can know that any instance of the following is true when the blanks are replaced by two occurrences of the same declarative sentence of my language:

'.....' in my language means that ..... [from (1b) and (1c)]

(1f) I can know that '"S is W" in my language means that S is W' is a declarative sentence of my language [from (1b) and (1c)]

(1g) I can know that 'S is W' is a declarative sentence of my language [from (1b), (1c) and (1f)]

(1a) I can know that '"S is W" in my language means that S is W' is true [from (1e)

and (1g)].

The only premiss that now looks at all vulnerable is (2b). But *if* philosophical reflection can establish particular instances of semantic externalism as conceptual truths then (2b) (or some substitute which will serve just as well for the argument) is safe, so it looks as if the sceptic must argue that philosophical reflection cannot establish particular instances of externalism as conceptual truths. So he must either find fault with the arguments for (particular instances of) externalism of Putnam, Burge and so on, or argue that these are arguments the soundness of which cannot be established by philosophical reflection. (The sceptic may object that the philosophical reflection in question can only be carried out by someone who has the concepts of snow, whiteness and BIVhood, which the BIV has not, so it is question-begging to appeal to premiss (2b) – and indeed, even (2a) – in the context of this debate. But this is not so for a reason emphasised several times already: the basis of the sceptic's claim is that I cannot know that I am not a BIV *even if* I am not, so it is legitimate in responding to him to *assume* that I have the conceptual resources that I have if I am not.)

However that may be, if (1) is acceptable the following argument would seem to be sound:

(1) I can know that 'S is W' in my language means that S is W [from (1b) and (1c)]

(1h) I can know that if 'S is W' in my language means that S is W then I am capable of entertaining the thought that S is W

(1i) I can know that I am capable of entertaining the thought that S is W [from (1) and (1h)]

Since this pattern of argument can be repeated for any declarative sentence of my language that I understand we have the general conclusion that: *for any declarative sentence of my language that I understand, I can know of the thought that sentence in fact expresses that I am capable of entertaining that thought.*

It follows that anyone inclined to regard externalism as putting in question the possibility of knowledge of the content of one's (externalistically determined) thoughts must say which of (1b), (1c) and (1h) it conflicts with, and how.

Finally, what is the significance of Putnam's argument? It is a familiar point that the brain in the vat hypothesis it refutes is a very specific one; it does not, for example, bear on the hypothesis that I am a recently envatted BIV. But what the argument shows is that given

that instances of semantic externalism can be established as a conceptual truths the sceptic's claim that his sceptical hypothesis cannot be known to be false even if it is false – because it cannot be known to be false *a priori* and cannot be known to be false *a posteriori* if it cannot be known to be false *a priori* – is not universally true. And so the sceptic needs an argument to convince us that it is ever true.

References

Putnam, H. 1981: *Reason, Truth and History*. Cambridge: Cambridge University Press.

Wright, C. 1992: 'On Putnam's proof that we are not brains-in-a-vat'. *Proceedings of the Aristotelian Society*, vol. LXXXXII, pp. 67-94.